Conducting experiments on cultural aspects of document design: Why and how?

HANS HOEKEN and HUBERT KORZILIUS

Abstract

In order to answer the question whether it is wise to adapt a document to the culture it is to be used in, one can conduct an experiment with 'cultural difference' as one of the main variables. In this article, we discuss three problems that researchers encounter when conducting such experiments. First, employing 'nationality' to operationalize cultural differences leads to interpretation problems when differences in responses occurs. Cultures differ from each other on a large number of dimensions. Each dimension constitutes an alternative explanation for any difference in response between members from different cultures. Second, it is difficult to construct documents and measurement instruments that are equivalent in all cultures included in a survey. The question is whether documents can have equivalent meanings, and whether questionnaires measure the same concept in two or more cultures. Third, members of certain cultures are reluctant to use the ends of rating scales, whereas members of other cultures use them freely. For each of these problems, we present solutions.

Keywords: cultural differences, document design, experiment, 'extremity of response' problem, China, United States, product documentation

Introduction

Virtually every handbook on marketing states that companies have to grow in order to attain long-term stability. A company that is very successful on its home market may have to export its business to other countries to generate more business. Whether it will be successful in these other countries depends on several factors: the extent to which its products provide value for money compared to its competitors, the efficiency with which the company can distribute its product, and the

Communications 28 (2003), 285–304 03412059/2003/028–0285

quality of its communication efforts. We are especially interested in this last factor.

Communicating effectively is difficult enough when a company operates in its home market, but when operating in another country it becomes even more difficult because cultural differences may inhibit smooth and effective communication. Most studies about problems in intercultural communication have focused on problems in interpersonal communication. However, in the next section, we argue that problems arising as a result of cultural differences in document design are important as well.

In this article, we do not present an empirical study on cultural differences in document design. Instead, we present an overview of the problems one encounters when conducting such studies. Of course, methodological problems in studying cultural differences have received attention by other scholars (see, e. g., Brown and Sechrest, 1980; van de Vijver and Leung, 1997, 2000), but none of these scholars focuses on the specific problems that arise when conducting experiments including the use of documents. We try to bring together both the problems and solutions in one article. We illustrate these problems and solutions by referring to studies that have been conducted on cultural differences in document design.

Why study cultural differences in document design?

Research on issues in intercultural communication has focused mainly on interpersonal communication (see, e. g., Lustig and Koester, 1999; Scollon and Scollon, 1995; Wiseman and Koester, 1993). This interest may be generated by the direct feedback people get when things go wrong in interpersonal communication; their conversation partner may look puzzled, start to laugh, or feel insulted as a result of a seemingly inoffensive remark. Producers of advertisements and manuals lack such direct feedback.[1] Only when their mistakes become too apparent, do they end up in David Ricks's collection of blunders in international business (Ricks, 1999).

The success of a company depends partly on the quality of its documents. This is, for example, true for the product manuals a company provides to help a consumer to use the product as efficiently as possible. When a user get frustrated because the manual lacks quality, he or she may become dissatisfied with the product. Studies in the US (Schriver, 1997) and the Netherlands (Jansen and Balijon, 2002) have shown that consumers are willing to pay more for a product when it is accompanied by high-quality documentation. Furthermore, these same studies show that people are more inclined to buy another product by the same company after a favorable experience with a product's documentation.

A company's success also depends on the quality of its marketing communication. Such communication is designed to inform consumers about the availability and attributes of a company's products, to create a positive attitude toward the product and toward the company, and to convince consumers to buy the product. Several studies have shown that, in order to attain these goals, it is important to adapt these communications to the culture the expressions of communication are aimed at (see e. g., De Mooij, 1998).

Fiske, Kitayama, Markus and Nesbitt (1998) note that cultural differences can be studied and described in a number of ways. A successful approach has been to classify cultures according to differences in their value hierarchies (see for a review, Smith and Schwartz, 1997). What is considered to be an important value in one culture (e. g., status), might be considered relatively unimportant in another. Based on results from one of the largest of these studies, Hofstede (1984) claims that such differences in value hierarchies form the core of cultural differences.

The fact that cultural differences reside in differences in value hierarchies will have important repercussions for the effectiveness of marketing communication, as values are an essential element in the persuasion process. One way to convince consumers to buy a company's products is to argue that the product is of a superior quality. To that end, an advertisement or direct mail should convince consumers that the product's attributes provide them with important benefits. The relation between a product's attributes and benefits is called a means-end-value chain (Reynolds and Gutman, 1988). The product's attributes are the means that enable the user to attain certain benefits (= the ends). The extent to which users appreciate these benefits depends on their values.

Companies that try to sell their products using a product's attributes as an argument should reflect on the relative importance of values in a culture. Take, for instance, the values of safety and status, and a company that produces an expensive car that distinguishes itself from its competitors on these two attributes. First, the car is safer than most of its competitors. Second, its high price signals the high social status of the owners. Marketing communication often has difficulty enough getting one point across, let alone two. Therefore, the company has to decide whether to stress the car's safety or status. The former strategy is probably more effective in a culture in which safety is considered more important than status, whereas the latter strategy may be more effective in a culture in which the hierarchical order of these values is reversed.

Apart from promoting a product by stressing its attributes, companies try to gain a competitive advantage using communication differently, namely linking their brand to a certain value. One reason for this strategy may be that the product does not distinguish itself in any other

aspect from its competitors; another reason may be that consumers are unwilling (or unable) to undertake the effort of evaluating a product's attributes and benefits. Most people have difficulty distinguishing one brand of cigarettes from another when they are ignorant of the cigarette brand they are smoking. That is, the cigarette brands do not differ much from one another with respect to their functional qualities. To distinguish themselves from their competitors, companies try to add an extra value to their brand through communication. For instance, some companies try to build a strong association between their brand and adventure. Others add the value of being part of the jet set, whereas still others try to attach the value of independence to their brand. Given that the importance of values differs from one culture to another, adding a certain value to one's product may be more successful in one culture than in the other.

Research on cultural differences in document design

Cultural differences can be relevant to all types of documents. Studies have documented cultural differences in the design of such diverse documents as memos and letters (Bell, Dillon and Becker, 1995), teaching contracts (Stevens, 2000), risk information (Sauer, 1996), direct mail letters (Graves, 1997) and business writing in general (Tebeaux, 1999). However, the majority of studies on cultural differences in document design assess the influence of culture on the design of advertisements.

Most of the latter studies are content-analytic in nature. The researchers collect a large number of advertisements that have been published in different countries, and they analyze these ads with respect to the information they contain, the strategy used (hard sell or soft sell), the type of values that are appealed to, and the kind of images that are used. Although the quality of these studies is sometimes questionable (see Harris and Attour, 2000, for a discussion), there are clear indications that cultural differences in advertising strategies exist. In a carefully conducted meta-analysis of 59 content-analytic studies, Abernethy and Franke (1996) conclude that cultural differences affect the amount of information included in advertisements. Le Pair, Crijns, and Hoeken (2000) reviewed 17 content-analytic studies that discuss a total of 145 comparisons of characteristics of advertisements from different cultures. More than 60% of these comparisons were statistically significant. Therefore, both studies indicate that there are indeed differences between cultures in the design of advertisements.

The results of these studies show that document designers from different cultural backgrounds differ in their opinions on what makes a document effective.[2] Sauer (1996), for instance, reports differences be-

tween English and U.S. documents designed to provide risk and safety information for mineworkers. Apparently, English document designers differed from their American colleagues in their ideas on the effectiveness of several rhetorical choices. However, it might be that the English documentation would have been equally or even more effective for the American mineworkers compared to the American documentation (or the other way around: the American documentation may prove to be most effective for both countries.) To test these possibilities, an experiment should be conducted in which the two documents are evaluated by mineworkers from both countries. In the next section, we show that conducting such experiments and interpreting their results can be quite difficult.

Problems in conducting (cross- and intercultural) experiments

In this article, we discuss three problems in conducting experiments in which cultural difference is employed as an independent variable. These problems are derived from more general discussions of problems in conducting research on intercultural issues (Brown and Sechrest, 1980; van de Vijver and Leung, 1997). To illustrate these problems, we discuss an experiment reported on by Zhang and Gelb (1996). Their study focused on the persuasiveness of different advertising appeals in different cultures.

Zhang and Gelb (1996) created advertisements to promote a photo camera. In one version, the slogan read "Come and indulge in the joy of self-expression", the other read "Share the moments of happiness with your family and friends". The former appeal was expected to be more congruent with an individualistic culture such as the US; the latter appeal was expected be more congruent with a collectivistic culture such as the People's Republic of China (PRC). Zhang and Gelb created both an English and a Chinese version of each advertisement. They, thereupon, had American students (from a university in the northeastern part of the United States) and Chinese students (from a university in southern China) read one version of the advertisement. Participants were asked to indicate what they thought of the product advertised and their opinion on the advertisement itself.

Zhang and Gelb hypothesized that a congruent appeal would be more persuasive than an incongruent appeal. This indeed proved to be the case. The self-expression appeal was congruent and more persuasive for the (individualistic) US participants; the sharing with family and friends appeal was congruent and more persuasive for the (collectivistic) Chinese participants. Therefore, this study provides evidence to support the statement of adapting an advertisement to appeal to the dominant values in a certain culture.

We use the study by Zhang and Gelb (1996) to illustrate the issues researchers encounter when conducting such experiments. The first issue deals with interpreting the interaction between nationality (Chinese or American) and the type of advertising appeal. The second issue concerns the translation of the English advertisements and questionnaires into Chinese. This may have resulted in (subtle) differences in meaning. Finally, cultural differences in applying the extremes on answering scales are important. That is, participants from some cultures are less inclined to tick extreme responses than participants from other cultures. This is called the "extremity of response problem". These three problems are discussed along with possible solutions. Zhang and Gelb (1996) are aware of these issues and enable us to show solutions to these problems.

The invalid inference problem

When designing an experiment, researchers want to be able to conclude that an observed difference on the dependent variable (in this case, persuasiveness) can only be caused by the variation on the independent variables (in this case, culture and appeal). Zhang and Gelb (1996) operationalized cultural background by nationality (American versus Chinese), which leads us to our first issue. Nationality hosts a number of differences between American and Chinese participants other than the cultural differences in value hierarchies. The PRC and US differ, for example, with respect to their history, legal and political systems, and economic situation. Each of these factors provides a rival explanation for an effect observed on the dependent variable. For instance, advertising has only been allowed in the PRC since 1978. Therefore, Chinese advertisers as well as consumers have much less experience than American advertisers and consumers in producing and interpreting advertisements. Such a difference in experience may have made the Chinese participants more susceptible to certain appeals than the Americans. On the basis of the data reported by Zhang and Gelb, this explanation cannot be ruled out.[3]

Several researchers have suggested that this problem can be solved by finding correlates of the cultural differences at the individual level (see e. g., Singelis and Brown, 1995; Van de Vijver and Leung, 1997). Cultural differences are often described at group level (see e. g., Hofstede, 1984). These descriptions are based on the responses of individuals from that culture, but they are not, as Hofstede himself warns, an adequate description of each and every individual. For example, the fact that the US as a group scores relatively high on the masculinity dimension in the Hofstede study does not imply that each and every American scores high

on this dimension. American society has its 'Rambos' as well as its 'Woody Allens'. Likewise, a description of the Chinese culture value hierarchy should not be mistaken to be a description of the value hierarchy of each and every Chinese person. However, as a result of socialization processes, the value hierarchy of an individual Chinese citizen is more likely to correspond roughly to the Chinese culture's hierarchy than the value hierarchy of an individual American citizen would.

Zhang and Gelb hypothesize that the Chinese participants are more susceptible to the sharing with friends and family appeal in their advertisement because the value harmony takes a higher position in the Chinese value hierarchy than in the American value hierarchy. Likewise, the American participants are more susceptible to appeals of self-expression as the value 'individual freedom' occupies a higher position in the American value hierarchy than in the Chinese. Information on the value hierarchy of individual participants would have enabled us to test this hypothesis more stringently. First, we could have checked whether the majority of the American participants indeed thought individual freedom more important than harmony, and whether the opposite could be said for the majority of Chinese participants. Second, we could have tested whether the 'self-expression' appeal was more convincing for people estimating individual freedom higher than harmony, and whether the 'sharing with friends and relatives' appeal was more convincing to people who value harmony more than individual freedom, regardless of their nationality. If these two tests were to be carried out, we could be more certain in our conclusion that cultural differences in value hierarchies lead to differences in persuasiveness of the two advertising appeals.

The logic behind this strategy of using individual correlates is as follows. Zhang and Gelb (1996) reason that the interaction between nationality and appeal occurs because the majority of the American participants attach more importance to individual freedom than to harmony, and, therefore, are more susceptible to the self-expression appeal. However, undoubtedly a limited number of American participants estimate harmony higher than individual freedom. If Zhang and Gelb are right, these participants will be more susceptible to the sharing with friends and relatives appeal. These dissenting American participants affect the difference in persuasiveness of the two appeals. If all American participants had valued individual freedom more than harmony, the difference would have been larger. The same line of thought applies to the Chinese participants. Although the majority attaches more importance to harmony than to individual freedom and, therefore, will yield more readily to the sharing with friends and relatives appeal than to the self-expression appeal, dissenters will hold the opposite view.

Should the interaction between nationality and appeal be the result of differing value hierarchies, then a comparison between the responses given by participants appraising 'individual value' over 'harmony' and the responses of participants preferring 'harmony' over 'individual value' would lead to a larger interaction effect, regardless of their nationality. In statistical terms: the amount of explained variance should increase. A smaller interaction effect indicates that the interaction between nationality and appeal is not caused by the difference in value hierarchy but by one of the other differences between American and Chinese participants.

Figures 1a through 1c provide (fictitious) graphic representations of this argument. Figure 1a represents the interaction between nationality (PRC, US) and appeal (friends and relatives, self-expression). It shows that the Chinese participants yield more to the sharing with friends and relatives appeal whereas the American participants yield more to the self-expression appeal.

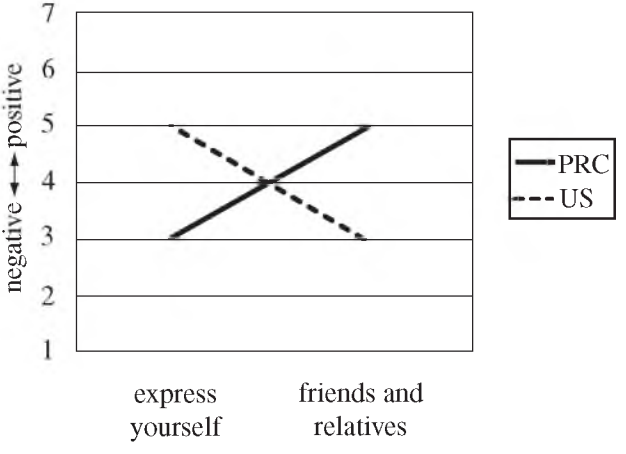

Figure 1a. *The attitude toward the brand as a function of nationality (PRC (People's Republic of China), US), and appeal (express yourself, friends and relatives) (1 = very negative, 7 = very positive).*

When the participants are grouped on their value hierarchy instead of on nationality, two results are possible. If a difference in value hierarchy leads to a difference in persuasiveness of the two appeals, the interaction between value hierarchy (harmony more important than personal freedom, personal freedom more important than harmony) and appeal (sharing with friends and relatives, self-expression) should become stronger. Figure 1b represents this scenario.

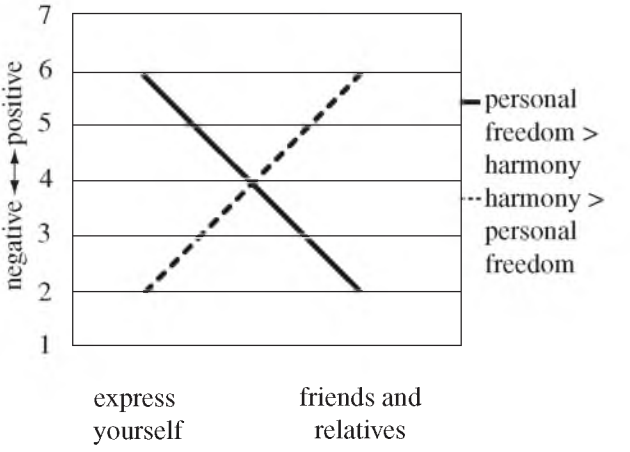

Figure 1b. *The attitude toward the brand as a function of personal value hierarchy ('personal freedom > harmony', 'harmony > personal freedom') and appeal (express yourself friends and relatives) if the difference in value hierarchy is the cause of the difference in response between the American and Chinese participants (1 = very negative, 7 = very positive).*

However, if the original interaction between nationality and appeal is caused by one of the other factors that co-vary with nationality (e. g., experience with commercial advertising), the interaction between value hierarchy and appeal may become weaker (see Figure 1c). As a result of grouping by value hierarchy, the group which values harmony more than personal freedom contains American participants who have abundant experience with commercial advertising; likewise, the group that values personal freedom more contains Chinese participants who have little experience with commercial advertising. These dissenters more or less pollute the effect of experience with commercial advertising, thereby reducing the interaction between appeal and value hierarchy.

The following procedure can be applied to test statistically whether the interaction between nationality and appeal is caused by the difference in value hierarchies or by another factor. Using multiple regression, two interaction terms can be defined: Appeal x Nationality and Appeal x Value hierarchy. These two interaction terms can be entered into the analysis. If the difference in Value hierarchy is indeed the explanatory factor, the interaction between Appeal x Value hierarchy is significant whereas the Appeal x Nationality interaction is non-significant. If the Appeal x Nationality interaction is significant while the Appeal x Value hierarchy is non-significant, this interaction must be caused by another factor than the hypothesized cultural difference. A clear illustration of

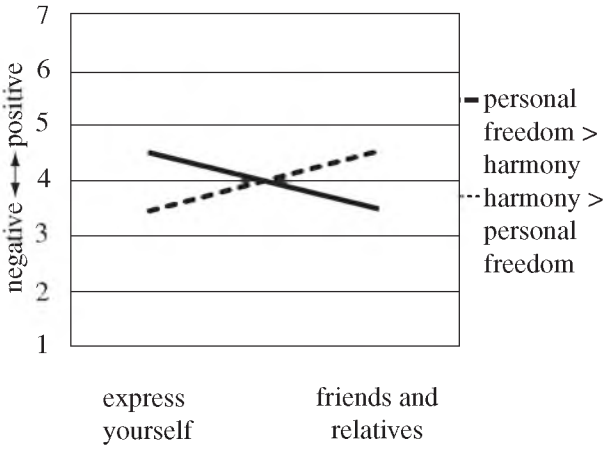

Figure 1c. *The attitude toward the brand as a function of personal value hierarchy ('personal freedom > harmony', 'harmony > personal freedom') and appeal (express yourself friends and relatives) if the difference in value hierarchy is NOT the cause of the difference in response between the American and Chinese participants (1 = very negative, 7 = very positive).*

this strategy is the study conducted by Wang, Bristol, Mowen and Chakraborty (2000).

In summary, the problem in using nationality as the operationalization of cultural differences is that nations differ from each other on a number of dimensions. Each of these differences may be responsible for any difference obtained. A solution to this problem is to measure an individual characteristic of the participants that is related to the cultural difference at the group level. To researchers interested in the difference in value hierarchies, the value list developed by Schwartz (1992) may prove useful. This value list has been developed and used extensively in many countries and languages. The individual scores on the value list can be used to divide participants into two groups; one that regards one value as more important than the other (regardless of nationality) and another group for which the opposite is true. If the cultural difference is indeed responsible for the difference obtained, then a grouping based on the scores on this individual characteristic yields larger differences than a grouping based on nationality. This solution presupposes the existence of a measurement instrument that is equivalent in the different cultural groups. In the next section on the second problem in conducting these experiments, we argue that constructing equivalent documents and measurement instruments is easier said than done.

The problem of document and measurement instrument equivalence

In cross-cultural experiments on document design, different versions of a document and the items for the questionnaire are usually developed in one language, and then translated into the other language included in the experiment. The assumption is that the document carries the same meaning in both languages. This assumption is often not warranted.

Constructing equivalent documents may prove difficult for several reasons. Certain concepts are culture-specific and lack an equivalent in another culture. For instance, the Dutch concept of 'gezelligheid', which constitutes a blend of togetherness and enjoyment, is difficult to translate in another language. Providing an equivalent translation may be difficult, especially when culture-related concepts are at stake. Most researchers are aware of this issue, but they may not be aware of the fact that the illustrations used in documents may lend themselves to different interpretations as well. Messaris (1997, p. 109) provides an excellent example of this phenomenon. He reports on a study in which US and Chinese participants were asked for their interpretation of a Fisher-Price ad in which a man in jeans explains a toy to a little boy. The U.S. participants thought this man to be a member of the upper-middle class (e. g., "He is probably a yuppie, like the rest of us"), whereas the Chinese participants thought the man was working class. Half of the participants explained their interpretation by referring to the fact that the man was wearing jeans (which they considered to be only suitable as working clothes for a manual laborer). This example shows that even identical illustrations may carry different meanings for participants from different cultures.

Equivalence problems do not only arise in the context of documents, they can also arise in the development of measurement instruments (see Brislin, 1986, pp. 143–150, for advice on how to word items). In studies in which participants evaluate the quality of a document, responses are often collected using semantic differentials and Likert-items. The connotations of words used in these items may differ between cultures. Even when different cultures appear to share the same language, differences in connotations may occur. Although Dutch is spoken in both the Netherlands and Flanders, for instance, problems in meaning equivalence may arise (see Scott, 2000, for a discussion of similar problems in American and British vocabulary).

For instance, when one is interested in the effects of a corporate image advertising campaign in Flanders and the Netherlands, one may ask participants to indicate to what extent they regard the company as 'voortvarend' (= energetic). If the campaign is successful, the responses given in the Netherlands may vary strongly from the responses given in

Flanders. The Dutch participants may indicate that they agree with this statement, whereas the Flemish participants may indicate that they disagree with it. This difference is not the result of a cultural difference in response to the campaign, but rather a result of a difference in connotation. In the Netherlands 'voortvarend' carries the meaning of 'acting boldly', which has a positive connotation, in contrast to Flanders, where 'voortvarend' carries the meaning of 'acting hurriedly'. When such problems in measurement equivalence arise even when two cultures share the same lexicon, this type of problem may be even worse when items have to be translated.

How can equivalence problems be detected? Zhang and Gelb (1996) used the "translation-back translation" method recommended by Brislin (1980) to construct equivalent documents. Specifically, the English version was translated into Chinese by one translator. Another translator translated the Chinese version into English. This procedure yielded two English versions of the same ad. The more these two versions carry the same meaning, the more certain one can be that the Chinese version also carries the same meaning. When the two English versions differ from each other in important respects, this may be an indication that the Chinese version is not equivalent. However, other explanations may also be possible. The difference between the two English versions may be the result of a flawed translation from the original English version into Chinese (resulting in two non-equivalent versions), or it may be the result of a flawed back translation from the Chinese version into English. Discussion between the translators could help in order to determine the source of the confusion. The 'translation – back translation' method can also be employed to develop approximately equivalent measurement instruments.

Another method is suggested by Erkut, Alarcón, Coll, Tropp and Vazquez (1999). Instead of a translation-driven approach, they recommend using a concept-driven approach. In this approach, native speakers of the different languages independently develop an instrument to measure the same concepts. After the instrument is developed in two languages, the resulting items are compared, and the items are rephrased to be as similar as possible with respect to factors such as grammatical form.

Once the study has been conducted, statistical methods can be used to test whether the measurement instruments have been equivalent. Especially when several items are used to measure a concept, the correlations between these items can indicate whether the measurement instrument was equivalent in both cultures. To illustrate this option, let us suppose that we have conducted a study on the effectiveness of a corporate image campaign in the Netherlands and Flanders in which we measured the dynamic nature of the company on three items: I find the company 'daadkrachtig' (i. e., decisive), I find the company 'actief' (i. e.,

active), and I find the company 'voortvarend' (i. e., 'bold' in The Netherlands, 'rash' in Flanders).

These three qualifications all carry positive connotations in the Netherlands, but in Flanders 'voortvarend' carries a negative connotation. When comparing the correlations between the scores on the items for the Dutch and the Flemish participants separately, the following picture would arise (Figure 2).

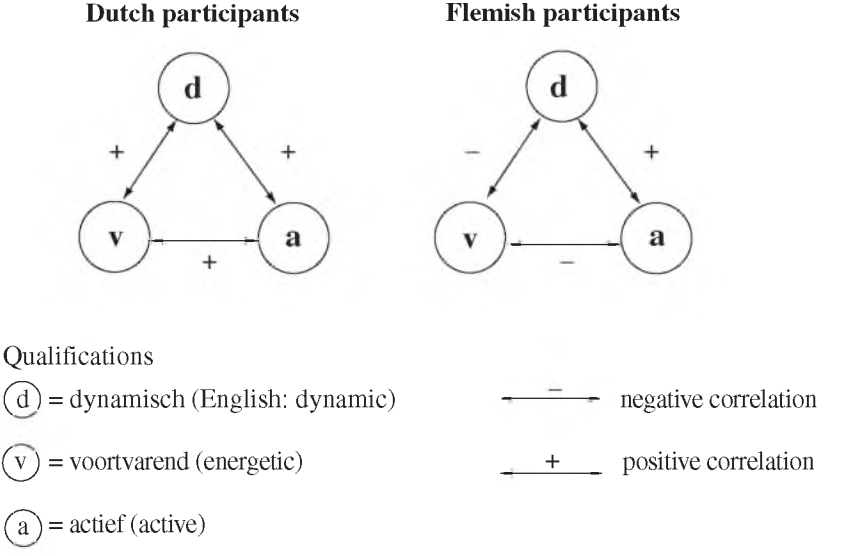

Figure 2. *Different correlations between identically worded items because Dutch and Flemish participants have different connotations (Dutch: voortvarend = bold; Flemish: voortvarend = rash).*

A positive response to the 'daadkrachtig'-item would in general be followed by a positive response to the 'voortvarend'-item for the Dutch participants, but by a negative response to the 'voortvarend'-item for the Flemish participants. That is, for the Dutch participants a positive correlation between the responses on the two items would be obtained whereas for the Flemish participants a negative correlation between the items would arise. Therefore, the internal-consistency of the measurement instrument would be higher for the Dutch participants than for the Flemish participants. Computing Cronbach's alpha is usually recommended to establish the internal-consistency reliability of a measurement instrument consisting of several items. This coefficient takes the corre-

lations between the items into account, and, as a result, Cronbach's alpha for the Dutch version of the measurement instrument differs from that for the Flemish version of the measurement instrument. A difference in the reliability of the measurement instrument is, therefore, a signal that the instruments are not equivalent (see Van de Vijver and Leung, 1997, for a review of relevant statistical techniques). However, even if one is successful in constructing equivalent measurement instruments, cultural differences in the use of these instruments may yield a problem.

Extremity of response problem

In a typical experiment, participants are required to respond on a five- or seven-point scale. For instance, their intention to buy the advertised product is measured by asking the respondents to indicate the extent to which they agree with the statement "I would definitely buy this product" on a scale ranging from "strongly disagree" to "strongly agree".

Several problems related to cultural differences may arise when using this method. For instance, participants from different cultures may be more or less experienced in filling out such questionnaires, which may influence their responses. Furthermore, when numbers are used, as is often the case, participants with different cultural backgrounds may have different associations with these numbers. For instance, in the Dutch educational system, grades range from 1 (lowest grade) to 10 (highest grade). In the German educational system grades 1 to 6 are used but in Germany low numbers indicate high grades. This difference between the two countries may be reflected in response behavior. A partial solution to this problem is not to label scale points with numbers, but to provide only verbal labels (Krosnick and Fabrigar, 1997).

A critical issue is constituted by cultural differences in response style. Respondents from certain cultures are more likely to use the extremes on an answering scale (that is, ticking 'strongly disagree' or 'strongly agree'), than respondents from other cultures, who appear to avoid these extremes. De Mooij (1998, p. 132) reports that when 5% of Japanese respondents indicate that they will definitely buy the product, the product will be a huge success. On the other hand, when 55% of Italian respondents state that they will definitely buy the product, it will fail. These results indicate that Japanese respondents are less likely to use the ends of the scale than the Italians do.

Zhang and Gelb (1996) compared the responses of Chinese and American participants. They were aware that Chinese participants are less likely to use the end points of the scale than U.S. participants (Chun, Campbell and Yoo, 1974). This difference in response style may lead to faulty conclusions. Figure 3a represents the interaction between nationality and appeal.

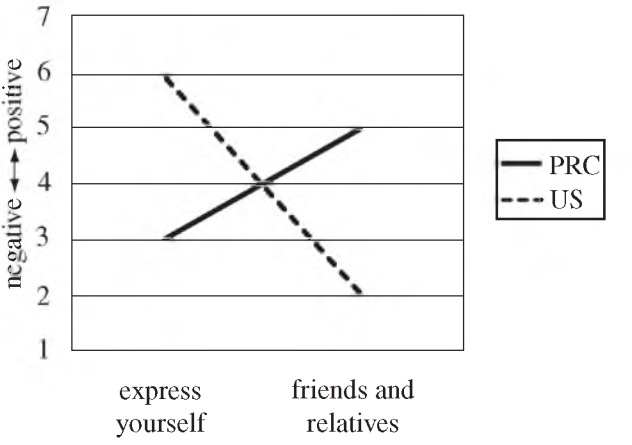

Figure 3a. *The interaction between nationality (People's Republic China, PRC, versus US), and appeal (express yourself versus friends and relatives) for the attitude toward the brand suggesting a stronger appeal effect for the American participants (1 = very negative, 7 = very positive).*

This (fictitious) figure suggests that the difference in persuasiveness between the sharing with friends and relatives appeal and the self-expression appeal is much larger for the American participants than for the Chinese participants. This may indeed be the case, but this pattern of results may also be caused by the difference in response style. Whereas the American participants may have freely used the extremes of the scale, resulting in means of 6.00 and 2.00, the Chinese participants may have refrained from using the extremes, resulting in means of 3.00 and 5.00. To test whether the difference in effect is a real or an artificial effect as a result of response style differences, one must compare the analysis of the raw data with the analysis of the standardized data (cf. Van de Vijver and Leung, 1997).

The logic of this operation is as follows. If Chinese participants do not (frequently) use the end points of a seven-point scale, they will use only the numbers 2 through 6. The American participants on the other hand, will use the numbers 1 through 7 more freely. As a result, the standard deviation of the Chinese sample is lower than the standard deviation of the US sample. When using standardized scores for each country separately, this difference in standard deviation disappears (because all standardized scores have a standard deviation of 1). If the standardized scores are analyzed in the same way as the raw scores, two outcomes are possible.

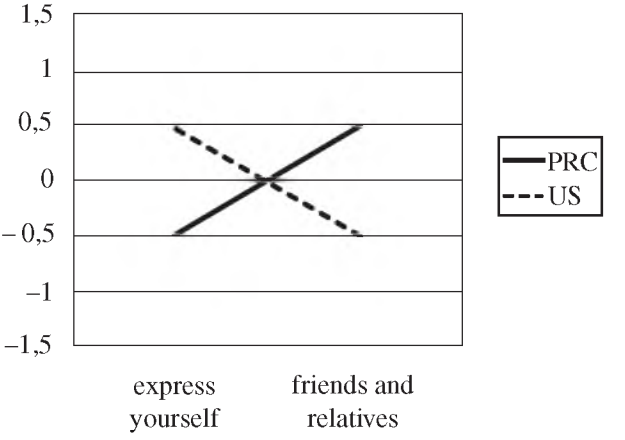

Figure 3b. *The interaction between nationality (People's Republic China, PRC, versus US), and appeal (express yourself versus friends and relatives) for the attitude toward the brand if the stronger appeal effect for the American participants is caused by the differences in the use of extremes (units are standard scores, the deviation of a specific score from the mean expressed in standard deviation units, ranging from negative to positive attitudes toward the brand).*

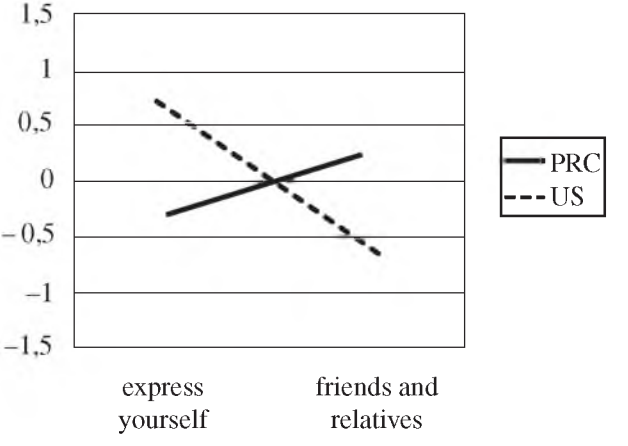

Figure 3c. *The interaction between nationality (People's Republic China, PRC, versus US), and appeal (express yourself versus friends and relatives) for the attitude toward the brand if the stronger appeal effect for the American participants is NOT caused by the differences in the use of extremes (units are standard scores, the deviation of a specific score from the mean expressed in standard deviation units, ranging from negative to positive attitudes toward the brand).*

If the difference in effect is indeed caused by the Chinese participants refraining from using the scale's extremes, this difference will disappear when standardized scores are used (see Figure 3b).

However, if the difference in effect is a genuine effect, showing that the effect of different appeals is indeed smaller for the Chinese participants than for the American participants, this difference in effect will reappear in the analysis of standardized scores (see Figure 3c).

Zhang and Gelb (1996) reported only standardized scores in their study.

Conclusion

Most studies on cultural differences in document design use content analysis as their research method. Many of these studies report significant differences between design characteristics from different cultures. These differences show that designers from different cultures differ in respect to their intuitions about what makes a document convincing. However, these intuitions may be incorrect. To study whether people with a different background are susceptible to different design characteristics, experiments should be conducted. Relatively few such studies have been conducted. This should not come as a surprise if one realizes the problems researchers are confronted with when conducting such experiments.

We have discussed three major problems in this area of research. We illustrated these problems using an experiment addressing the question whether appeals in advertisements should be adapted to the dominant values of a culture. However, the problems we have discussed are relevant when conducting experiments using other types of documents as well. First, employing nationality to operationalize cultural differences leads to problems in the interpretation of any difference in response. Cultures differ from each other on a large number of dimensions. Each of these dimensions poses an alternative explanation for any difference in response between members from different cultures. The solution to this problem is to measure an individual characteristic that is related to the cultural difference. Scores at this level make it possible to test whether a difference in response is related to the cultural difference or whether it should be ascribed to one of the other differences between the cultures.

Second, constructing documents and measurement instruments that are equivalent in both cultures is difficult. The 'translation-back translation' method can be used to assess the degree to which the document and instruments are equivalent. Third, there are cultural differences in the employment of rating scales. Members of certain cultures are reluc-

tant to use the ends of the scales, whereas members of other cultures do not feel inhibited to do so. Statistical techniques can be used to correct for differences in extreme response style. We have tried to show that studying inter- and cross-cultural aspects of document design is important, challenging, and, although difficult, very much possible.

Notes

1. An exception has to be made for feedback provided by reader-focused document evaluations methods (Schriver, 1989). However, in recent reviews of the applicability of such methods, no attention is paid to the potential of these methods to detect cultural differences in the comprehensibility and acceptability of documents (De Jong and Schellens, 2000, 2001).
2. It is good practice in advertising to have a sample of the target audience evaluate the advertisements before they are published. Therefore, one could argue that the advertisers' intuitions are corroborated by the responses of the audience. In these type of tests, however, the audience is usually confronted with only one or a limited number of versions of the advertisement. Consequently, one does not know whether they would respond even more favorably to a radically different type of advertisement.
3. Using nationality only to operationalize cultural differences was very common during the period in which Zhang and Gelb conducted their experiment (Segall, Dasen, Berry and Poortinga, 1999: 217–218). We could have used other studies to illustrate this point, for instance, Han and Shavitt (1994).

References

Abernethy, A. M. and Franke, G. R. (1996). The information content of advertising: A meta-analysis. *Journal of Advertising, 25(Summer)*, 1–17.

Bell, A. H., Dillon, W. T. and Harald Becker, H. (1995). German memo and letter style. *Journal of Business and Technical Documentation*, *9*, 219–227.

Brislin, R. W. (1980). Translation and content analysis of oral and written materials. In Triandis, H. C. and Berry, J. W. (Eds.), *Handbook of cross-cultural psychology. Methodology* (*Vol. 2,* pp. 389–344). Boston, MA: Allyn and Bacon.

Brislin, R. W. (1986). The wording and translation of research instruments. In Lonner, W. J. and Berry, J. W. (Eds.), *Field methods in cross-cultural research* (pp. 137–164). Newbury Park, CA: Sage.

Brown, E. D. and Sechrest, L. (1980). Experiments in cross-cultural research. In Triandis, H. C. and Berry, J. W. (Eds.), *Handbook of cross-cultural psychology. Methodology* (*Vol. 2*, pp. 297–318). Boston, MA: Allyn and Bacon.

Chun, K., Campbell, J. C. and Yoo, Y. H. (1974). Extreme response style in cross-cultural research. *Journal of Cross-Cultural Psychology, 5,* 465–480.

Erkut, S., Alarcón, O., Coll, C. G., Tropp, L. R. and Vázquez García, H. A. (1999). The dual-focus approach to creating bilingual measures. *Journal of Cross-Cultural Psychology, 30,* 206–18.

Fiske, A. P., Kitayama, S., Markus, H. R. and Nisbett, R. E. (1998). The cultural matrix of social psychology. In Gilbert, D. T., Fiske, S. T. and Lindzey, G. (Eds.), *The handbook of social psychology* (4th Ed., Vol. II, pp. 915–981). New York: McGraw-Hill.

Graves, R. (1997). 'Dear Friend'(?): Culture and genre in American and Canadian direct marketing letters. *Journal of Business Communication, 34,* 235–252.

Han, S. and Shavitt, S. (1994). Persuasion and culture: Advertising appeals in individualistic and collectivistic societies. *Journal of Experimental Social Psychology, 30,* 326–350.

Harris, G. and Attour, S. (2000). Content analysis of advertising 1970–1997: A review and assessment of methodologies. In Monye, S. O. (Ed.), *The handbook of international marketing communications* (pp. 237–249). Oxford: Blackwell.

Hofstede, G. (1984). *Culture's consequences.* Newbury Park, CA: Sage.

Jong, M. D. T., de and Schellens, P. J. (2000). Toward a document evaluation methodology: What does research tell us about the validity and reliability of evaluation methods? *IEEE Transactions on Professional Communication, 43,* 242–260.

Jong, M. D. T. de and Schellens, P. J. (2001). Optimizing public information brochures. Formative evaluation in document design processes. In Janssen, D. and Neutelings, R. (Eds.), *Reading and writing public documents* (pp. 59–83). Amsterdam: Benjamins.

Jansen, C. and Balijon, S. (2002). How do Dutch consumers use instruction guides? Confirming and disconfirming patterns of use. *Document Design, 3,* 194–204.

Lustig, M. W. and Koester, J. (1999). *Intercultural competence. Interpersonal communication across cultures (3rd Ed.).* New York: Longman.

Messaris, P. (1997). *Visual persuasion. The role of images in advertising.* Thousand Oaks, CA: Sage.

Mooij, M. de. (1998). *Global marketing and advertising.* Thousand Oaks, CA: Sage.

Krosnick, J. A. and Fabrigar, L. R. (1997). Designing rating scales for effective measurement in surveys. In Lyberg, L. E., Biemer, P., Collins, M., De Leeuw, E., Dippo, E., Schwarz, N. and Trewin, D. (Eds.), *Survey, measurement and process quality* (pp. 141–164). New York, NY: Wiley.

Pair, R. le, Crijns, R. and Hoeken, H. (2000). Het belang van cultuurverschillen voor het ontwerp van persuasieve teksten. [The importance of cultural differences for the design of persuasive texts.] *Tijdschrift voor Taalbeheersing, 22,* 358–372.

Reynolds, T. J. and Gutman, J. (1988). Laddering theory, method, analysis, and interpretation. *Journal of Advertising Research, 28,* 11–31.

Ricks, D. A. (1999) *Blunders in international business (3rd. Ed.).* Oxford: Blackwell.

Sauer, B. A. (1996). Communicating risk in a cross-cultural context. A cross-cultural comparison of rhetorical and social understanding in U.S. and British mine and safety training programs. *Journal of Business and Technical Communication, 10,* 306–329.

Schriver, K. A. (1989). Evaluating text quality: The continuum from text-focused to reader-focused methods. *IEEE Transactions on Professional Communication, 32,* 238–255.

Schriver, K. A. (1997). *Dynamics in Document Design.* New York: Wiley.

Schwartz, S. H. (1992) Universals in the content and structure of values: theoretical advances and empirical tests in 20 countries. In Zanna, M. P. (Ed.), *Advances in Experimental Social Psychology* (Vol. *25,* pp. 1–65). San Diego, CA: Academic Press.

Scollon, R. and S. W. Scollon (1995). *Intercultural communication. A discourse approach.* Oxford: Blackwell.

Scott, J. C. (2000). Differences in American and British vocabulary: Implications for international business communication. *Business Communication Quarterly, 63,* 27–39.

Segall, M. H., Dasen, P. R., Berry, J. W. and Poortinga, Y. H. (1999). *Human behavior in global perspective. An introduction to cross-cultural psychology (2nd Ed.).* Needham Heights, MA: Allyn and Bacon.

Singelis, T. M. and Brown, W. J. (1995). Culture, self, and collectivist communication. Linking culture to individual behavior. *Human Communication Research, 21,* 354–389.

Smith, P. B. and Schwartz, S. H. (1997). Values. In Berry, J. W., Segall, M. H. and Kagitcibasi, C. (Eds.), *Handbook of Cross-Cultural Psychology. Volume 3: Social behavior and applications* (pp. 77–118). Boston, MA: Allyn and Bacon.

Stevens, B. (2000). Russian teaching contracts. An examination of cultural influence and genre. *Journal of Business and Technical Communication, 14,* 38–57.

Tebeaux, E. (1999). Designing written business communication along the shifting cultural continuum. The new face of Mexico. *Journal of Business and Technical Communication, 13,* 49–85.

Vijver, F. van de and Leung, K. (1997). *Methods and data analysis for cross-cultural research.* Thousand Oaks, CA: Sage.

Vijver, F. van de and Leung, K. (2000). Methodological issues in psychological research in culture. *Journal of Cross-Cultural Psychology, 31,* 33–51.

Wang, C. L., Bristol, T., Mowen, J. C. and Chakraborty, G. (2000). Alternative modes of self-construal: Dimensions of connectedness-separateness and advertising appeals to the cultural and gender-specific self. *Journal of Consumer Psychology, 9,* 107–15.

Wiseman, R. L. and Koester, J. (Eds.). (1993). *Intercultural communication competence*. Newbury Park, CA: Sage.

Zhang, Y. and Gelb, B. D. (1996). Matching advertising appeals to culture: The influence of products' use conditions. *Journal of Advertising, 25,* 29–46.